IJBST

ISSN: 0974 – 3987
IJBST (2010), 3(4):46-54

Comparative genomics and phylogenetic study of *Plasmodium falciparum* and *Plasmodium vivax*

Manoj Kumar Yadav[1], Alok Mishra[2] and D. Swati[3 *]

1. Department of Bioinformatics, MMV, Banaras Hindu University, Varanasi-221005, INDIA
2. Department of Physics, MMV, Banaras Hindu University, Varanasi-221005, INDIA
3. Department of Bioinformatics and Physics, MMV, Banaras Hindu University, Varanasi-221005, INDIA
swatid@gmail.com

ABSTRACT
Malaria is a lethal human disease in terms of morbidity and mortality and is caused by *Plasmodium* species. Main obstacle in conquering this disease is their rapidly evolving genetic structure. Recently published whole genome of *Plasmodium vivax* provides an opportunity to compare it in depth with the previously published genome of *Plasmodium falciparum*. Here we have compared each and every chromosome of both the species and found the surprising presence of an unknown bases N. By analyzing tandem repeats it was clear that its copy number is always high in case of *P.falciparum* and it has strong correlation with their chromosome length. All the *vir* and *var* genes are also analyzed and very little sequence homology found between them. In order to find the functional role of *vir* and *var* gene protein products, amino acid composition is analyzed. The frequency of hydrophilic amino acids is found more in comparison with the hydrophobic one. Phylogenetic analysis of mitochondrial cytochrome b and nuclear encoded High mobility group binding (HMGB) proteins was done. On the basis of phylogeny result of HMGB it was found that this transcription factor can act as a reference point for the development of separate drugs for *P.falciparum* and *P. vivax* and this strategy is supposed to be more reliable in the eradication of malaria.

Key words: Allelic diversity, phylogeny, tandem repeats, vertical transfer, invasiveness.

INTRODUCTION

Malaria is one of the most common infectious disease and an enormous health problem. Each year, up to three million deaths due to malaria and close to five billion episodes of clinical illness possibly meriting anti-malarial therapy occur throughout the world, with Africa having more than 90% of this burden[1,2]. Almost 3% of disability adjusted life years are due to malaria mortality globally. Malaria occurs in the widespread parts of the Americas, Asia, and Africa. Ninety percent of malaria-related deaths occur in Sub-Saharan Africa. Malaria is commonly associated with poverty, but is also a cause of poverty and a major hindrance to economic development [1, 3].

Malaria is caused by a number of species of *Plasmodium*, among them the *Plasmodium falciparum* is the deadliest one [4]. This most virulent form causes malaria by invading both the reticulocytes and mature erythrocytes, while it's another species *Plasmodium vivax* is less virulent form, restricted to only reticulocytes [5]. From the starting a lot of research is going on in this field to eradicate the disease but our scientific community is little bit successful in their efforts. One of the main reasons behind this low successful rate is the rapidly evolving genetic structure of *Plasmodium* parasite itself. A lot of drugs were made to combat this disease but maximum of them fails in their initial clinical trial phases. This mainly happens due to development of resistance against anti malarial drugs. Till date no effective vaccines are not available for malaria [6,7,8], and also the available drugs in the market usually become less effective due to the high rate of antigenic variations [9].

Recently the whole genome sequence information on *Plasmodium vivax* became available [10]. On combining this information with the already available whole genome information on one of the deadliest species, *Plasmodium falciparum* [11] have opened up exciting research avenues in analyzing their pathogenic role at the genomic level. Because of fast throughput sequencing methods, a new era of *in silico* analysis started which can help in studying both the genomes comprehensively. In this study we focused on all the 14 chromosome of *Plasmodium falciparum* and *Plasmodium vivax* species and did their comparative studies of two genes which are mainly responsible for high rate of antigenic variations. The main differences in the genome of the two species lie in the Adenine and Thymine (A+T) composition, *P.vivax* has approximately 55% and *P.falciparum* has approximately 80%, being the second highest AT-rich genome [4, 12].

Among a number of virulence factors responsible for antigenic variation, the Erythrocyte Membrane Protein1 is most prominent one. This variant protein is encoded by *var* genes in the whole genome [3, 13, 14]. These *var* genes are mostly found in the telomeric and sub-telomeric regions; exceptions are chromosome 4, 7, 8 and 12 in which they specifically lie in the central region [www.vardb.org, 15, 16].

IJBST

ISSN: 0974 – 3987
IJBST (2010), 3(4):46-54

Phylogenetic analysis of *Plasmodium* species is well studied previously for revealing the taxonomical status of the apicomplexan parasite. For this analysis the highly conserved mitochondrial genomes or proteins are taken into account due to their vertical transfer by means of maternal inheritance. Despite of these facts we proposed a new strategy of phylogenetic analysis based on medium level of sequence conservation of transcription factors for the sake of designing more potent vaccine against malaria.

MATERIALS AND METHODS

DNA sequence of all the 14 chromosomes of *P.falciparum* and *P.vivax* were collected from the Gene bank in NCBI (www.ncbi.nlm.nih.gov). *var* genes that are responsible for the antigenic variation in *P.falciparum* are taken from the PlasmoDB (www.plasmodb.org/plasmo/). This is the database that contains all the proteomic and genomic information regarding the *P.falciparum.* The percent GC and AT content of all the chromosome of both the species were calculated by writing a program with the help of MATLAB version 7.4 (www.mathworks.com). We have selected all the *var* genes that are scattered throughout the genome of *P.falciparum.* To study the *P.vivax* genome, all the *vir* genes are taken from the NCBI database (www.ncbi.nlm.nih.gov). To check the sequence similarity among the foresaid genes, we have used the Clustal W [17].

In order to detect the tandem repeats in the whole genome of *P.vivax* and *P.falciparum*, we used the Tandem Repeats finder [18]. This software is used for comparing the frequency of occurrence of tandem repeats in each and every chromosome of both the species on the basis of their sequences, lengths, and copy numbers. Correlation analysis and curve fitting of tandem repeats have been done [19, 20]. Percent compositions of each of the four nucleotides are also calculated by using the same.

The protein sequences that are encoded by *var* and *vir* genes were obtained from NCBI (www.ncbi.nlm.nih.gov/nucleotide). The amino acid composition of these protein sequences was detected by using the MEGA 4.0 [21, 22].

We downloaded the highly conserved Cytochrome b mitochondrial protein of different *Plasmodium* species. Next we selected High mobility group binding protein (HMGB) which is one of the most conserved transcription factors in *Plasmodium* species [23]. Both the Cytochrome b and HMGB were extracted from the NCBI database

(www.ncbi.nlm.nih.gov/ protein). In order to find the similar sequence in different *Plasmodium* species, we used the PSI-BLAST program [24]. Complete sequences were aligned in CLUSTAL W using the default parameters. We performed the Maximum parsimony method for constructing the phylogenetic tree. Maximum parsimony analysis was performed on MEGA 4.0 [21, 22]. A total of 100000 replicates were carried out for the bootstrap analysis.

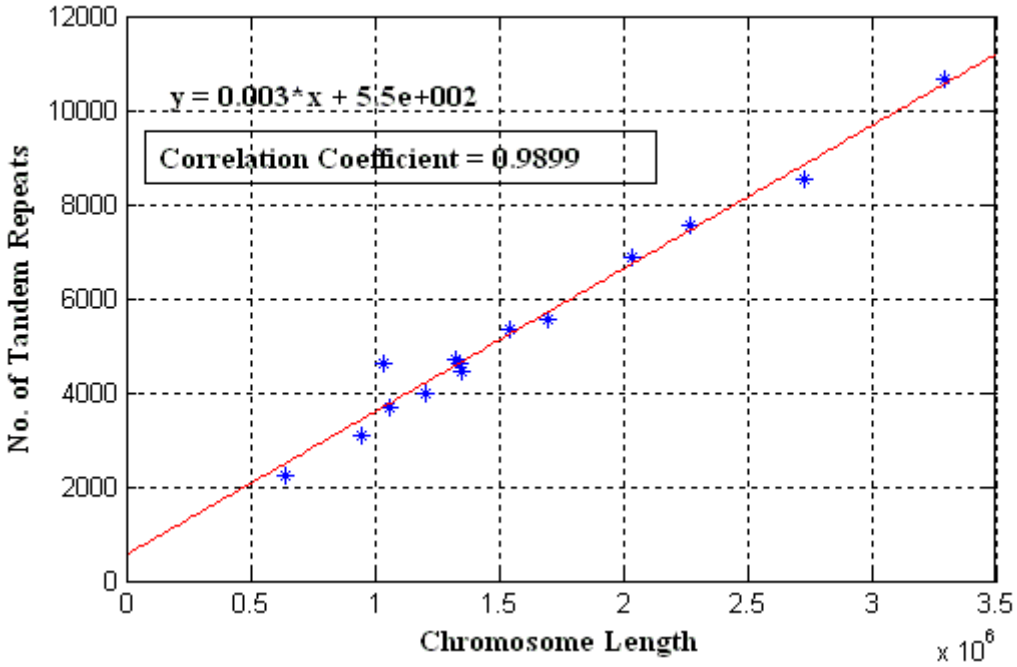

Figure1. (a) Number of tandem repeats vs. chromosome length and the Best Fit for *P.falciparum*

ISSN: 0974 – 3987
IJBST (2010), 3(4):46-54

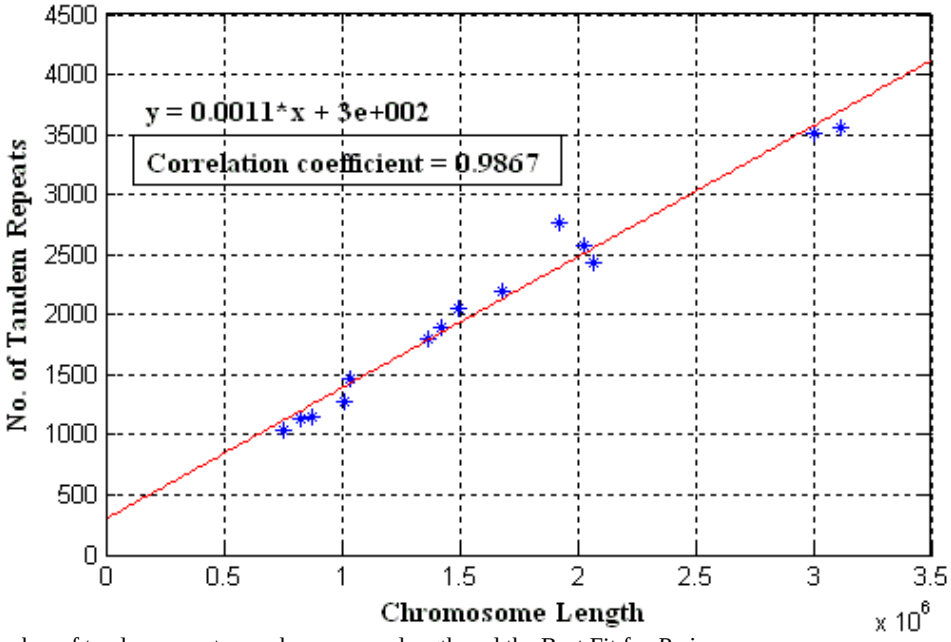

Figure1. (b) Number of tandem repeats vs. chromosome length and the Best Fit for *P.vivax.*

RESULTS

A comparison of whole genome sequence information of each and every chromosome of *P.falciparum* and *P.vivax* (Table 1) reveals several interesting features of the two respective genomes. Although both the species have same number of chromosomes, their genome size is different.

The genome size of *P. vivax* is larger but when we goes through analyzing each chromosome, we find that the length of chromosome number 2, 3, 4, 6, 10, 13 and 14 is larger in *P.falciparum*. Comparing A+T content in both the species by our program gave some results that are different from those quoted in NCBI (www.ncbi.nlm.nih.gov/protein). This is basically due to occurrence of some mysterious bases N's found apart from A, T, C, and G bases.

Analyzing the *vir* and *var* gene in *P.vivax* and *P.falciparum* respectively, we find very little sequence similarity. The mean length of *var* genes are also more when compared with the *vir* one.

In order to identify the tandem repeats in each chromosome, we found that numbers of tandem repeats are relatively high in case of *P.falciparum* [Table1]. Tandem repeats are showing strong correlation with the chromosome length in both the species [Figure 1].In case of *P.falciparum* the total number of repeats was 75,952, while in *P.vivax* its number was 28,817. For finding tandem repeats in *vir* and *var* gene, we apply similar strategy and found that *P.falciparum var* genes are found to be much longer in length and contains 724 tandem repeats in comparison with that of *P.vivax vir* genes, which contain 14 tandem repeats.

All the protein sequences encoded by *vir* and *var* genes were analyzed. Very little similarity was found among the *var* gene products and *vir* gene products also. The amino acid composition is very important in determining the functional role of these proteins [25] .So overall amino acid composition is calculated with the help of MEGA 4.0 [21, 22] and found that most of the *var* and *vir* gene products contains gluatamate, lysine, asparagine and serine in large proportion [Table 2, 3].

The normalized average frequency of each hydrophilic amino acid is higher for Vir and Var proteins than it should be in case of unbiased statistical distribution of amino acids.

To study the taxonomical status of *P.falciparum* and *P.vivax*, the phylogenetic tree is reconstructed with the help of more conserved cytochrome b mitochondrial protein (Figure 1). This tree (Figure 3 a) shows the taxonomical position of both the species *vis-a-vis* each other. Among the various transcription factors we have used High mobility groupbinding protein (HMGB) for analyzing phylogenetic relationship due to their high level of conservation. The phylogenetic tree reconstructed by taking HMGB shows a distant relationship among the *P.falciparum* and *P.vivax*

IJBST

ISSN: 0974 – 3987
IJBST (2010), 3(4):46-54

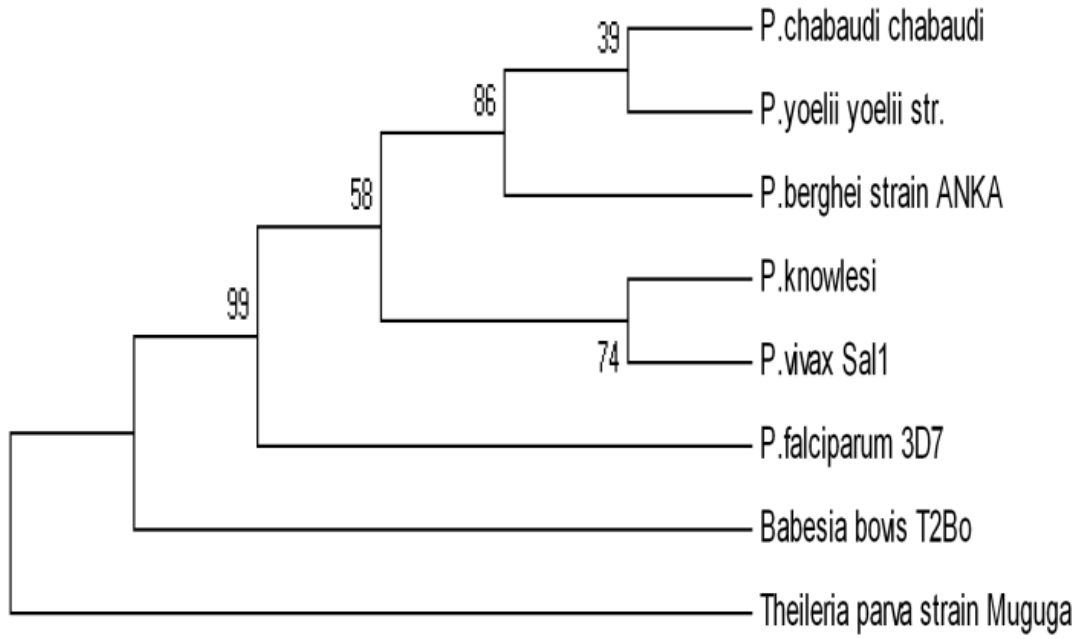

Figure 3. (a) Phylogenetic tree derived from an alignment of selected Cytochrome b mitochondrial protein.

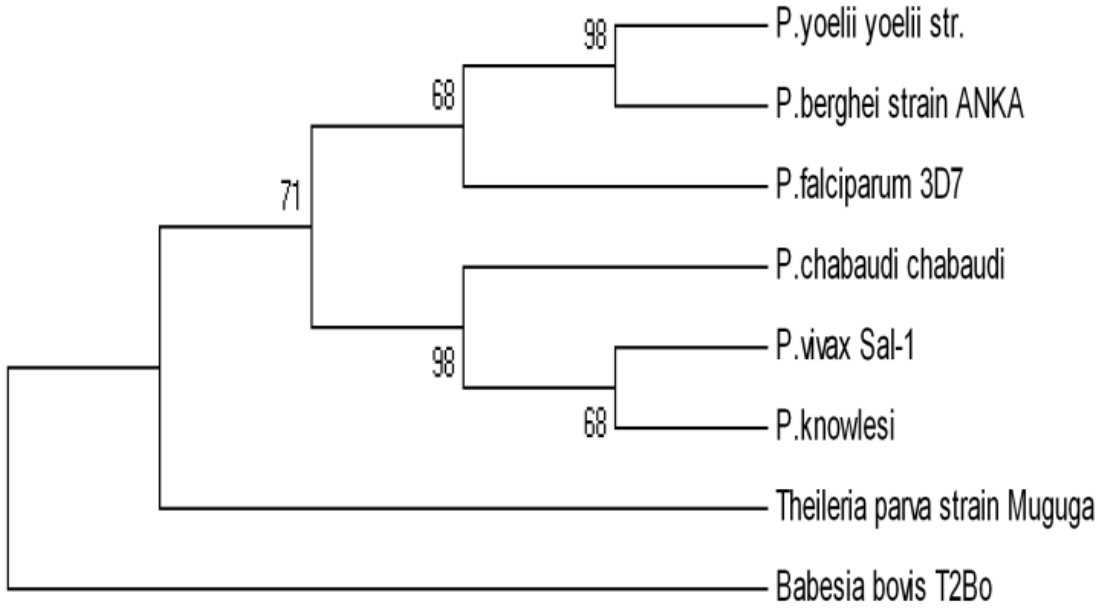

Figure 3. (b) Phylogenetic tree derived from an alignment of selected nuclear encoded High mobility group binding protein.

IJBST

ISSN: 0974 – 3987
IJBST (2010), 3(4):46-54

Table 1. Comparative analysis of various genomic features of *P. vivax* and *P. falciparum*

Chromosome Number	Characters	*P.vivax*	*P.falciparum*
1	Size (bp)	830,022	643,292
	% AT	52.88	79.45
	No. of Genes	176	157
	No. of Tandem repeats	1,129	2,221
2	Size (bp)	755,035	947,102
	% AT	55.13	80.25
	No. of Genes	162	223
	No. of Tandem repeats	1,043	3,106
3	Size (bp)	1,011,127	1,060,087
	% AT	55.18	80.12
	No. of Genes	220	247
	No. of Tandem repeats		3,674
4	Size(bp)	876,622	1,204,112
	% AT	54.86	79.32
	No. of Genes	208	254
	No. of Tandem repeats		4,000
5	Size (bp)	1,370,936	1,343,552
	% AT	55.72	80.67
	No. of Genes	316	330
	No. of Tandem repeats	1,806	4,632
6	Size(bp)	1,033,388	1,418,244
	% AT	54.08	80.21
	No. of Genes	248	322
	No. of Tandem repeats	1,468	4,635
7	Size(bp)	1,497,819	1,351,552
	% AT	54.37	80.02
	No. of Genes	353	294
	No. of repeats	2,047	4,458
8	Size (bp)	1,678,596	1,325,595
	% AT	54.60	80.29
	No. of Genes	378	299
	No. of Tandem repeats		4,690
9	Size (bp)	1,923,364	1,541,723
	% AT	53.91	80.98
	No. of Genes	434	367
	No. of Tandem repeats		5,353
10	Size (bp)	1,419,739	1,694,445
	% AT	55.05	80.31
	No. of Genes	327	404
	No. of Tandem repeats	1,897	5,572
11	Size (bp)	2,067,354	2,035,250
	% AT	54.92	81.04
	No. of Genes	468	490
	No. of Tandem repeats	2,429	6,886
12	Size (bp)	3,004,884	2,271,916
	% AT	55.37	80.69
	No. of Genes	695	530
	No. of Tandem repeats		7,550
13	Size (bp)	2,031,768	2,732,359
	% AT	54.35	80.85
	No. of Genes	446	667
	No. of Tandem repeats		8,535
14	Size(bp)	3,120,417	3,291,006
	% AT	56.99	81.56
	No. of Genes	699	771
	No. of Tandem repeats	3,558	10,640

IJBST

ISSN: 0974 – 3987
IJBST (2010), 3(4):46-54

Table 2. Amino Acid composition of *vir* gene encoded proteins in *Plasmodium vivax*

Accession No.	Different types of Amino Acids composition																			
	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
XP 001608611.1	4.51	3.07	6.35	9.63	4.3	6.76	1.43	5.74	8.4	6.56	1.64	9.43	3.48	1.23	3.89	8.81	3.69	5.12	0.41	5.53
XP 001608613.1	3.23	2.82	7.46	9.48	4.23	2.62	2.22	7.26	12.1	5.04	1.61	6.45	3.02	2.42	6.65	5.85	4.23	5.24	0.81	7.26
XP 001613646.1	3.13	4.11	7.57	7.4	6.41	2.96	1.48	7.24	11.68	7.57	1.97	9.54	1.97	3.29	2.96	6.58	2.8	4.11	0.99	6.25
XP 001612659.1	1.58	3.95	5.93	7.71	5.14	2.77	3.75	6.32	12.25	10.28	1.78	8.3	2.37	1.98	2.77	6.92	4.15	3.36	0.79	7.91
XP 001612596.1	6.11	1.81	4.75	7.01	4.52	6.56	0.9	5.2	12.44	6.11	0.9	5.2	4.52	3.85	2.94	10.41	4.07	5.66	1.13	5.88
XP 001612578.1	2.89	3.07	6.15	7.59	3.62	4.88	2.89	6.87	11.75	6.33	2.17	9.22	2.35	1.45	4.88	8.68	4.88	4.34	0.72	5.24
XP 001612579.1	1.11	3.62	3.9	9.19	6.41	4.46	1.95	6.96	11.98	5.57	1.67	9.19	1.95	1.67	4.74	8.91	2.79	3.62	1.11	9.19
XP 001612835.1	1.95	3.41	6.01	9.09	6.17	2.76	1.95	6.33	11.69	7.79	1.62	10.06	3.41	3.08	1.95	7.63	3.73	3.73	0.81	6.82
XP 001612837.1	4.09	2.59	4.53	8.62	5.17	3.66	1.72	7.33	12.72	6.03	0.43	8.19	4.53	1.29	2.16	8.84	6.9	3.88	0.65	6.68
XP 001612937.1	1.31	3.76	7.35	7.03	5.88	1.96	2.94	7.35	12.09	8.01	0.98	10.46	2.94	2.61	1.31	6.37	4.74	3.76	0.98	8.17
XP 001612938.1	2.05	4.57	7.56	8.03	5.04	2.2	2.52	8.19	12.28	7.4	1.1	8.35	2.52	1.89	2.99	5.83	5.67	2.83	0.63	8.35
XP 001612941.1	1.55	3.75	6.4	9.27	2.87	3.09	2.65	5.96	11.26	6.62	1.1	8.61	3.53	1.32	3.97	9.05	6.4	5.3	1.55	5.74
XP 001608558.1	7.07	2.49	4.37	12.06	3.12	4.37	2.7	2.49	10.4	7.07	1.25	4.99	7.48	3.12	2.29	6.65	5.41	7.9	0.62	4.16
XP 001614849.1	2.24	3.81	5.61	8.3	4.93	5.16	0.9	7.62	10.99	8.3	1.35	11.88	2.69	1.79	2.69	7.4	4.71	3.59	0.22	5.83
XP 001614850.1	2.23	2.87	6.69	6.69	7.96	3.82	1.59	6.37	9.24	6.37	2.23	8.28	5.1	1.91	4.46	6.05	4.78	4.46	0.64	8.28
XP 001614853.1	2.51	3.63	6.7	7.54	6.42	3.63	2.79	6.15	8.1	9.22	1.68	8.94	4.19	2.51	3.35	5.87	6.98	3.91	1.12	4.75
XP 001614854.1	4.47	3.63	6.15	6.98	5.03	3.07	2.23	6.15	8.66	7.54	1.68	8.94	3.35	3.63	4.75	5.59	3.91	4.47	1.4	8.38
XP 001615150.1	2.43	4.51	7.29	5.9	4.86	2.08	2.43	4.86	10.07	12.15	3.13	7.99	2.43	1.04	1.39	8.33	4.51	4.86	0.69	9.03
XP 001615151.1	2.03	4.43	7.01	7.01	6.09	2.21	2.4	7.38	14.02	8.67	1.29	9.59	2.77	1.29	1.48	6.46	4.98	3.32	1.29	6.27
XP 001612686.1	3.94	4.26	6.47	7.26	4.73	3	2.37	6.47	12.15	8.83	2.21	8.99	2.21	2.21	2.68	6.15	3.79	4.89	0.63	6.78
XP 001612688.1	4.35	2.84	6.99	9.07	5.1	5.1	1.89	6.24	12.67	5.86	1.32	6.62	4.16	1.7	3.02	5.48	4.73	4.73	0.76	7.37
XP 001608384.1	5.58	3	6.44	6.44	3.43	6.87	2.15	5.36	10.94	5.79	1.29	6.44	3.65	4.08	2.79	9.23	6.44	5.36	0.64	4.08
XP 001608312.1	3.75	1.87	5.81	9.55	4.49	5.99	1.31	5.99	10.86	5.62	1.5	7.3	4.12	2.06	3.75	6.37	5.81	5.43	0.94	7.49
XP 001608316.1	3.61	2.77	6.8	7.21	4.16	5.83	2.64	3.88	8.6	8.32	1.39	6.8	3.19	3.47	5.69	7.77	5.41	3.88	1.11	7.49
XP 001608317.1	3.13	3.36	6.71	11.41	4.92	4.03	3.36	4.03	11.41	6.26	0.89	5.59	4.03	2.68	4.47	4.47	4.03	7.61	1.34	6.26
XP 001608318.1	3.47	3.82	7.47	8.51	4.86	2.6	2.26	7.47	12.33	7.99	0.52	8.16	2.26	2.26	2.6	6.94	5.21	3.99	0.69	6.6
XP 001608320.1	1.87	3.73	7.09	8.21	7.09	2.61	1.49	6.72	15.3	8.96	2.61	7.84	1.12	1.12	2.99	4.1	5.6	2.99	0.75	7.84
XP 001608323.1	2.53	4.52	7.23	9.04	5.97	3.44	1.81	7.05	13.92	6.33	1.45	7.78	2.35	1.81	2.35	5.06	3.44	4.16	0.54	9.22
XP 001614319.1	3.84	3.61	6.77	8.13	6.09	4.06	1.35	7	10.61	7.9	0.9	7.9	4.51	2.26	3.16	6.32	4.51	4.29	1.13	5.64
XP 001614320.1	8.22	1.2	4.97	9.25	3.6	5.99	2.23	2.74	11.64	6.68	0.86	5.31	6.85	3.08	2.05	8.05	5.48	6.34	0.68	4.79
XP 001614321.1	4.38	4.57	5.9	5.33	5.14	2.48	2.67	5.71	9.14	9.52	1.9	10.67	2.29	2.29	2.67	7.62	5.71	3.43	1.14	7.43
XP 001613040.1	3.33	5.19	6.11	7.41	5.93	2.22	1.85	6.85	11.11	8.52	1.85	10.74	2.22	2.04	2.41	6.11	4.44	3.15	0.56	7.96
XP 001613165.1	2.1	2.99	6.59	5.69	7.34	2.99	2.4	9.13	12.57	9.73	0.6	9.73	2.84	1.35	1.65	6.44	4.19	3.59	0.9	7.19
XP 001613166.1	2.32	4.11	6.61	5.89	3.21	1.79	2.5	7.32	12.5	9.11	1.96	10.54	3.04	1.96	2.14	7.32	7.32	3.04	0.71	6.61
XP 001616487.1	1.66	4.3	5.63	8.28	7.62	2.98	1.66	8.61	10.93	6.29	2.32	9.27	1.32	2.32	2.65	8.61	2.32	1.66	1.32	10.26
XP 001612629.1	2.5	2	5.5	8.5	4.75	6.25	1.75	5	7.75	8	1.5	6.25	3.75	2	6.75	9	5.25	4.75	0.75	8
XP 001612630.1	2.65	2.84	7.77	6.82	3.6	1.52	1.7	8.71	13.83	7.77	0.76	11.55	2.27	3.41	1.89	6.06	5.49	2.65	0.95	7.77
XP 001612631.1	1.65	3.3	8.07	8.26	4.22	3.3	1.83	6.06	11.01	7.89	1.83	7.52	3.12	2.39	3.3	10.28	4.95	3.85	0.37	6.79
XP 001612632.1	1.82	5	8.18	5.45	6.82	4.09	2.27	5.91	10.45	9.09	3.18	6.36	4.09	2.73	2.27	6.36	6.36	2.73	0.45	6.36
XP 001614314.1	1.57	3.83	7.65	9.04	5.04	3.3	2.09	7.65	14.09	9.74	0.87	8.87	2.78	1.39	1.57	5.74	3.3	2.78	1.22	7.48
XP 001614059.1	2.11	3.59	6.55	9.51	4.02	5.29	2.54	7.19	8.03	7.4	2.33	10.36	2.75	1.69	3.17	7.4	5.5	4.44	0.63	5.5
Avg	**3.14**	**3.46**	**6.49**	**8.04**	**5.04**	**3.75**	**2.16**	**6.44**	**11.41**	**7.66**	**1.48**	**8.46**	**3.24**	**2.26**	**3.07**	**7.09**	**4.83**	**4.23**	**0.85**	**6.9**

a) Number of hydrophilic amino acids is high.
b) All the data given here shows the percent composition.

IJBST

ISSN: 0974 – 3987
IJBST (2010), 3(4):46-54

Table 3. Amino Acid composition of *var* gene encoded proteins in *Plasmodium falciparum*

Accession No.	Different types of Amino Acids composition																			
	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
XP 001350936.1	4.72	3.33	8.55	8.18	2.64	6.2	1.66	4.76	11.05	5.13	1.29	7.17	5.46	4.39	3.19	5.78	6.84	3.88	1.62	4.16
XP 001351079.1	4.63	3.13	9.26	9.12	3.09	6.45	1.82	4.81	11.67	5.27	1.04	6.9	4.22	3.77	3.5	5.72	6.81	3.86	1.5	3.45
XP 001351080.1	3.94	3.08	8.28	9.51	2.94	6.75	1.54	4.75	11.18	5.61	0.86	7.42	5.34	3.62	3.67	6.29	5.98	4.39	1.36	3.49
XP 001351319.1	4.46	3.48	8.62	8.47	2.99	7.41	1.89	4.57	11.6	5.18	1.13	6.73	4.61	3.63	3.1	6.54	6.84	3.7	1.44	3.63
XP 001351321.1	4.04	3.35	8.16	9.09	2.86	6	1.67	4.85	11.36	5.62	1.41	8.13	5.13	3.29	4.01	5.86	5.6	3.61	1.76	4.21
XP 001351435.1	3.94	3.3	8.69	8.47	2.85	7.11	1.99	5.25	11.77	4.93	1.18	6.88	4.48	3.89	3.26	7.06	5.79	3.62	1.49	4.03
XP 001351437.1	4.66	3.03	8.48	8.3	2.59	7.77	1.84	4.44	10.5	5.93	1.49	7.42	4	3.95	3.51	6.54	6.59	3.91	1.19	3.86
XP 001351438.1	4.07	2.94	9.26	7.57	2.9	7.27	2.12	4.84	10.16	5.97	1.38	7.27	5.23	4.41	3.42	6.36	5.75	3.68	1.38	4.02
XP 001351439.1	4	2.92	9.2	7.82	2.92	7.18	2.19	4.9	10.19	5.85	1.25	7.22	5.25	4.47	3.35	6.28	5.72	3.83	1.38	4.08
XP 001351513.1	3.07	3.02	8.5	9.85	3.39	6.32	2.09	5.39	11.71	5.67	1.21	7.62	4.18	3.62	3.53	6.46	5.34	3.62	1.49	3.9
XP 001351514.1	3.41	3.04	8.5	9.5	3.04	6.77	2	4.45	12.13	5.86	1.09	7.63	4.45	3.23	3.54	6.54	5.72	4.18	1.41	3.5
XP 001351515.1	3.26	3.03	8.07	9.54	3.21	6.56	2.29	4.54	12.01	6.05	0.92	7.98	4.45	3.03	3.16	6.37	6.69	3.9	1.38	3.58
XP 001351517.1	4.1	3.28	8.8	8.89	3.1	7.07	1.73	4.65	11.49	5.2	1.23	7.39	4.1	3.83	3.15	6.16	6.61	3.92	1.5	3.78
XP 001351561.1	4.78	3.46	7.85	8.67	2.65	6.42	2	5.12	11.57	5.4	1.27	7.57	4.36	3.41	3.77	6.14	6.19	3.57	1.77	4.02
XP 001351564.1	3.72	3.13	7.52	10.33	2.63	6.21	1.54	4.98	11.01	5.48	1.04	6.62	5.21	4.08	3.44	6.62	7.39	3.67	1.4	3.99
XP 000965997.1	4.41	3.61	7.95	7.92	3.2	4.52	1.88	5.59	11.22	5.97	1.77	8.06	3.89	3.72	3.86	5.9	6.22	3.86	1.7	4.76
XP 000965999.1	3.4	2.49	8.08	7.17	2.79	4.83	2.04	6.34	9.74	6.11	2.72	11.4	3.62	2.72	4.08	6.49	4.91	4.83	1.58	4.68
XP 000966160.1	4.43	3.59	7.98	8.86	3.43	5.81	2.34	5.72	12.32	6.06	1.29	7.23	4.39	3.8	3.26	5.43	6.02	2.97	1.71	3.38
XP 000966305.1	3.89	3.62	8.3	8.22	3.26	5.21	2.05	5.54	11.73	5.87	1.39	7.76	4.05	3.74	3.31	6.5	5.77	3.26	1.77	4.75
XP 000966307.1	4.11	3.04	8.4	8.71	2.99	6.97	1.92	5.45	11.26	5.72	0.94	6.79	4.47	4.2	3.84	5.81	6.43	3.84	1.34	3.75
XP 001348946.1	4.78	3.46	7.85	8.67	2.65	6.42	2	5.12	11.57	5.4	1.27	7.57	4.36	3.41	3.77	6.14	6.19	3.57	1.77	4.02
XP 001349032.1	3.62	3.09	8.69	8.65	3	7.01	1.85	5.43	10.54	5.69	1.24	6.93	4.63	3.79	3.48	7.72	5.82	3.53	1.41	3.88
XP 001349035.1	3.44	2.95	8.31	8.71	2.99	6.16	2.64	5.94	10.67	5.58	1.25	8.22	4.87	4.11	3.44	6.25	6.16	2.99	1.43	3.89
XP 001349036.1	4.11	3.28	8.39	9.49	2.6	7.12	1.87	4.88	11.13	5.11	1.41	7.03	4.24	3.79	3.6	6.61	5.89	4.01	1.28	4.15
XP 001349219.1	3.58	3.39	7.86	8.67	3.08	6.55	1.81	5.32	11.95	5.82	0.89	7.4	4.01	3.16	3.08	6.78	7.44	4.01	1.31	3.89
XP 001349030.1	3.61	2.98	8.04	8.67	3.02	6.95	1.4	4.51	11.24	5.55	1.4	8.53	4.97	4.24	2.71	6.05	7	3.93	1.4	3.79
XP 001349031.1	4.04	2.86	8.62	9.36	3.38	6.15	1.98	5.05	11.52	5.41	1.27	7.43	4.13	4.18	3.6	6.46	5.71	3.43	1.32	4.09
XP 001349033.1	4.75	3.44	7.95	8.24	3.34	5.38	2.62	5.52	10.27	5.57	1.55	7.95	4.07	4.46	3.68	5.81	5.33	3.63	1.94	4.51
XP 001349034.1	4.42	3.21	8.49	9.16	3.21	6.77	1.9	5.15	11.38	5.73	1.44	7.45	4.24	3.57	3.16	5.73	6.14	3.88	1.35	3.61
XP 001349434.1	4.13	3.22	8.77	8.44	3.37	5.78	1.85	5.5	10.67	5.55	1.09	7.63	5.07	3.94	3.46	6.5	6.21	3.27	1.52	4.03
XP 001349437.1	3.77	3.14	8.96	10.49	2.91	6.32	1.93	4.57	12.33	4.98	1.48	7.31	3.99	3.72	3.45	5.65	6.1	3.81	1.34	3.77
XP 001349438.1	4.02	2.91	8.79	8.12	2.83	6.49	2.3	5.08	10.95	5.65	1.15	7.37	5.12	3.71	3.13	7.15	6.36	3.71	1.37	3.8
XP 001349512.1	4.43	3.59	7.75	8.83	3.02	5.64	1.81	5.17	11.31	5.57	1.11	7.82	5.03	3.62	3.05	6.48	5.77	3.93	1.64	4.43
XP 001349513.1	4.09	3.88	7.98	7.8	3.5	4.76	1.71	5.53	12.39	6.16	1.47	8.89	4.44	3.32	3.64	5.63	5.32	3.57	1.82	4.09
XP 001349514.1	4.1	3.21	8.96	8.77	2.97	6.04	1.75	5.14	10.8	5.71	0.99	6.6	5.38	4.06	3.35	6.37	6.27	4.01	1.46	4.06
XP 001351877.1	4.01	2.91	8.77	8.59	2.86	6.65	1.72	5.2	11.37	5.37	1.32	6.56	4.71	4.27	3.66	6.96	5.81	3.83	1.45	3.96
XP 001352242.1	3.71	2.92	8.97	9.14	2.74	6.54	1.77	5.26	11.66	5.65	1.37	6.93	4.2	3.45	3.62	6.36	6.63	3.53	1.41	4.15
XP 002585487.1	4.62	2.93	8.73	8.64	2.97	6.86	1.78	4.8	10.92	5.39	1.28	7.54	4.25	3.84	3.52	7.27	5.99	3.79	1.37	3.52
XP 001349738.1	3.7	3.1	7.96	9.9	2.68	5.92	1.67	5	11.05	5.18	1.34	7.12	5	3.75	3.33	6.01	7.03	4.39	1.48	4.39
XP 001350409.1	4.05	3.24	9.18	8.68	3.1	6.79	1.66	4.72	11.74	5.67	0.94	7.29	4.41	3.46	3.1	6.66	5.8	4.05	1.44	4
XP 001349740.1	3.26	3.68	8.19	9.41	3.23	5.71	1.79	5.29	11.78	5.68	1.49	7.53	4.66	3.56	3.77	6.28	5.26	3.89	1.76	3.8
Avg	**4.05**	**3.24**	**8.4**	**8.79**	**3**	**6.33**	**1.91**	**5.1**	**11.34**	**5.6**	**1.28**	**7.51**	**4.55**	**3.75**	**3.46**	**6.32**	**6.12**	**3.75**	**1.52**	**3.99**

a) Number of hydrophilic amino acids is high.

b) All the data given here shows the percent composition.

DISCUSSION

This study indicates that *Plasmodium vivax* and *Plasmodium falciparum*, both species are highly diverse at genomic level and this difference also reflects in their diverse functional role, as malaria parasites. The number of chromosome is same in both the species; *P.vivax* has a comparable larger genome size [10]. But there is remarkable difference in their chromosome lengths when considered separately. Some chromosomes of *P. vivax* are larger than the length of same chromosomes of *P. falciparum* and vice versa. However the major difference in their genome is their percent A+T content which is ~80 % in *P.falciparum*. Surprisingly, there is a significant distributional difference of A+ T nucleotides; as in *P.vivax* high A + T content is seen in the sub-telomeric regions [10], while in *P. falciparum*, the distribution is almost even in all the chromosome locations [11,26]. Since the A+T content is quite high in *P.falciparum* and considering high virulence property of the parasite, it may be possible that high A+T content will provide the base for remarkably high antigenic variation. This is because the fixation probability of alleles rich in A+T nucleotides are lower than G+C rich alleles and genes possessing antigenic and cyto-adherence properties maintain high allelic diversity without going to fixation and maintained by balancing selection [27]. A large number of N bases are distributed all over the chromosomes and their numbers goes on increasing as the chromosome length increase. We aim to analyze the role of the N's in context of genomic studies.

The length and nucleotide composition of *vir* and *var* genes are quite dissimilar. Hence their antigenic variation is also different and this may lead to difference in invasiveness. Generally the chromosome length is affected by the frequency of occurrence of tandem repeats. Overall numbers of tandem repeats are more in case of *P. falciparum* when compared with that of *P. vivax*. Although in both the species chromosome lengths are strongly correlated with tandem repeats and their correlation coefficients are nearly equal, but there is difference in the slope of the line of the best fit [Figure 1], which indicates that tandem repeats per unit length of chromosome is higher in *P.falciparum* in comparison to *P.vivax*. As we already know that the number of repeats is directly proportional to the relative simplicity factor (RSF), so high RSF value helps the organisms to avoid host's immune response [28, 29].

Although there is difference between the Phylogenetic tree of Cytochrome b and HMGB proteins, one point can be clearly predicted that both the species *P.falciparum* and *P.vivax* arose in a separate lineage from the common ancestor. Considering the HMGB tree, we can say that since HMGB is a very important transcription factor required for the survival of these species, hence by blocking the transcription process by some appropriate ligand blocking HMGB will halt the parasite's survival. This strategy will be helpful in drug development. The conservation level of HMGB proteins is very low among *P.falciparum* and *P.vivax* as shown in Phylogenetic tree (Fig.2 b), so a single drug will not work against both. Hence designing different drugs for *P.falciparum* and *P.vivax* may be a necessary strategy.

CONCLUSION

This study found many low complexity regions (LCR's) in *Plasmodium falciparum.* These LCR's have been shown to possess unfolded coil structures [30]. LCR are disordered regions comprising mostly of hydrophilic and low molecular weight amino acids, which impart flexibility with respect to folding and the resulting protein structure is therefore adaptable to ligand binding when required[31]. Natively unfolded/disordered regions in proteins confer considerable functional advantage as they allow efficient interaction with several different regions [32]. In context of possible drug targets, lower complexity regions, which were discarded before, will be taken into consideration and analyzed intensively in our next study.

REFERENCES:

[1] Breman JG. (2001) The intolerable burden of malaria: a new look at the numbers. Am J Trop Med Hyg *64* (Suppl 1−2):1–106.

[2] Sachs, J. & Malaney, P. (2002) The economic and social burden of malaria. Nature *415*, 680–685.

[3] Danquah et al. (2009) Reduced Efficacy of Intermittent Preventive Treatment of Malaria in Malnourished Children. Antimicrobial Agents and Chemotherapy. 53 (5): 1753.

[4] Dyer MD, Murali TM, Sobral BW. (2007) Computational prediction of host-pathogen protein protein interactions. ISMB/ECCB 23:159-166.

[5] Mendis K, Sina BJ, Marchesini P, Carter R. (2001) The neglected burden of *Plasmodium vivax* malaria. Am J Trop Med Hyg , 64**:**97- 106.

[6] Kyes SA, Kraemer SM, Smith JD. (2007) Antigenic variation in *Plasmodium falciparum*:gene organization and regulation of the var multigene family. Eukaryot Cell 6*:*11511-11520.

[7] Good, M and Kemp D. (2004) Overview of Vaccine Strategies for Malaria in New Generation Vaccines 3rd edition.

[8] Plassmeyer ML, Reiter K, Shimp RL, et al. (2009) The structure of the *Plasmodium falciparum*

circumsporozoite protein, a leading malaria vaccine candidate. J. Biol. Chem. 284 (39): 26951–63.

[9] Tinto H et al. (2006) In vitro susceptibility of *Plasmodium falciparum* to monodesethylamodiaquine, dihydroartemsinin and quinine in an area of high chloroquine resistance in Rwanda. Trans R soc Trop Med Hyg 100*:*509-514.

[10] Carlton, J. M. et al. (2008) Comparative genomics of the neglected human malaria parasite *Plasmodium vivax.* Nature 455, 757–763.

[11] Gardner, M. J. et al. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum.* Nature 419,498–511.

[12] N. Hall, M. Karras, J.D. Raine, J.M. Carlton, T.W. Kooij and M. Berriman et al. (2005) A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. Science 307, pp. 82–86.

[13] C.S. Janssen, R.S. Phillips, C.M. Turner and M.P. Barrett. (2004) *Plasmodium* interspersed repeats: the major multigene superfamily of malaria parasite. Nucleic Acids Res 32 , pp. 5712–5720.

[14] Martin R.E., Henry R.I., Abbey J.L., Clements J.D., Kirk K. (2005) The permeome of the malaria parasite: an overview of the membrane transport proteins of *Plasmodium falciparum.* Genome biology , 6: R26

[15] Hayes CN, Diez D, Joannin N, Honda W, Kanehisa M, Wahlgren M, Wheelock CE, Goto S. (2008) varDB: a pathogen-specific sequence database of protein families involved in antigenic variation. Bioinformatics (Oxford, England) 24:2564-5.

[16] Allred,D.R., A.F.Barbet, J.D. Barry, K. W. Deitsch. (2009) varDB: common ground for a shifting landscape. Trends Parasitol, 25 (56):249-52.

[17] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673-80.

[18] Benson G. (1999) Tandem Repeat finder: a program to analyze DNA sequences. Nucleic Acid Res 27:573-580.

[19] Zuluaga L, Pabón A., López C., Aleida Ochoa A., Blair *S. (2007*) Amodiaquine failure associated with erythrocytic glutathione in *Plasmodium falciparum* malaria. *Malaria Journal* , 6 *:* 47

[20] Coope, I.D. (1993) Circle fitting by linear and nonlinear least squares. Journal of Optimization Theory and Applications Volume *76*, Issue 2.

[21] Kumar S, Dudley J, Nei M & Tamura K. (2008) MEGA: A biologist –centric software for evolutionary analysis of DNA and protein sequences**.** Briefings in Bioinformatics 9: 299-306.

[22] Tamura K, Dudley J, Nei M & Kumar S. (*2007*) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24:1596-1599.

[23] Briquet S et al. (2006) High-Mobility-Group Box Nuclear Factors of *Plasmodium falciparum*. Eukaryotic cell p 672-682.

[24] Altschul, S.F *et al.* (1990) Basic local alignment search tool. J Mol Biology 215, 403–410.

[25] Hopp T. P. and Woods R.K. (1981) Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci., Vol. 78, No.6, pp.3824-3828.

[26] Barry, J. D., M. L. Ginger, P. Burton and R. McCulloch. (2003) Why are parasite contingency genes often associated with telomeres? Int J Parasitol, 33(1): 29–45.

[27] Bull PC, Buckee CO, Kyes S, Kortok MM, Thathy V, Guyah B, Stoute JA, Newbold CI, Marsh K. (2008) *Plasmodium falciparum* antigenic variation, Mappping mosaic var gene sequences onto a network of shared, highly polymorphic sequence blocks. Mol Microbiol 68:1519-1534.

[28] Hankock JM. (2002) Genome size and accumulation ofsimple sequence repeats: implications of new data from genome sequencing projects. Genetica 115:93-103.

[29] Vasee S Moorthy, W Ripley Ballou. (2009) Immunological mechanisms underlying protection mediated by RTS, S: a review of the available data**.** Malaria Journal, 8:312

[30] Elisabetta Pizzi and Clara Frontali (2001) Low complexity region in *Plasmodium falciparum* proteins. Genome Res. 11: 218-229.

[31] Ryan M Bannen, Craig A. Bingman, George N. Phillips Jr. (2008) Effect of low complexity regions on protein structure determination. J Struct Funct Genomics, 8:217–226.

[32] Liu,J., Tan,H. and Rost,B. (2002) Loopy proteins appear conserved in evolution. J. Mol. Biol., 322, 53–64.